

毕业设计任务书



题 目： 成都房产网络舆情分析平台设计—情感倾向性分析

学生姓名：

学 号：

专 业：

年 级：

学 院： 计算机与软件工程学院

指导教师：

**教务处制**

|  |
| --- |
| 毕业设计任务与要求：   1. **开发与运行** 2. 开发工具：C/C++/Python 3. 运行环境: Windows 4. **设计任务** 5. **开发团队**：**3人**。 6. **工作量**：代码量>1500行/人 7. **功能性**：利用成都房产网络舆情分析平台获取的各楼盘评论数据，进行情感倾向性分析。 8. **创新性**：用HMM模型和viterbi算法实现评论的情感词标注，以判断评论的情感倾向性，并给出对应楼盘的情感排名。 9. **实用性**：对不同楼盘评论情感倾向性的挖掘，有助于了解人们对楼盘的总体评价，开发商、物业等相关部门便于及时改变措施，提高服务水平。 10. **主要设计内容与要求**   主要内容包括对情感数据的分析，情感倾向性挖掘等，要使用目前比较流行的python,数据挖掘等进行相关的设计与实现。   * 1. 对爬取的评论数据进行分类（差评和好评）。   2. 通过好评的百分比实现楼盘评价排名。   3. 构建差评和好评的语料库。   4. 使用viterbi算法去通过HMM模型的相关参数对评论进行情感词性标记。   5. 通过评论的情感词性标记得到相关的正向评论和负向评论等。   要求对情感分析进行分析调研，然后分析归类，写出需求分析。   1. **本设计中的“复杂工程问题”** 2. **“复杂工程问题”的界定**   按照《工程教育专业认证标准》界定：复杂工程问题必须具备下述特征（1）；同时，具备下述特征（2）-（7）的部分或全部：   * 1. 必须运用深入的工程原理**，**经过分析才可能得到解决；   2. 涉及多方面的技术、工程和其他因素，并可能相互有一定冲突；   3. 需要通过建立合适的抽象模型才能解决，在建模过程中体现出创造性；   4. 不是仅靠常用方法就可以完全解决的；   5. 问题中涉及的因素可能没有完全包含在专业工程实践的标准和规范中；   6. 问题的相关各方利益不完全一致，存在利益冲突；   7. 具有较高的综合性，包含多个关联子问题。  1. **本设计具备“复杂工程问题”的特征**  * **具备特征（1）**：该分析平台需要对相关数据进行处理分析，要求现场调研、写出需求报告和可行性报告，故需要对开发的软件以工程方式进行。 * **具备特征（2）：**本软件涉及C/C++/Python技术，情感词性标记等。 * **具备特征（3）**：本软件需要进行现场调研，分析需求，建立前后台数据模型，需要用到数学知识和建模知识等。 * **具备特征（4）**：开发中要用到相关的模型与算法等。 * **具备特征（5）**：本系统有较高综合性，包括C/C++/Python知识的应用、HMM模型，viterbi算法等的使用。  1. **设计原则**   1. 独立开发：某些技术或算法可借鉴他人成果，但必须是在理解基础上的借鉴和改进，必须弄懂算法原理、技术手段。否则，视作抄袭，毕业设计成绩不合格。  2. 方案修改：与导师商议并取得导师同意后可适度修改,不得擅自删减。  3. 开发过程：遵循软件工程相关规定、阶段、步骤，熟悉工程化开发过程。   1. **相关文档撰写及要求** 2. 态度严谨，撰写认真，用词简洁准确，无语法、用词、术语、标点符号等错误。 3. 描述方式恰当，充分阐述设计理念、方案、算法、技术原理等，禁贴源代码代替。 4. 结构完整、层次清楚、图表美观、符合规定；图表必须绘制，禁止贴图。 5. 按学院学校对论文的要求认真撰写毕业设计说明书。 6. 按学院要求完成毕业设计任务的软件需求分析说明书。 7. 按学院要求完成毕业设计中所完成软件的测试工作及撰写测试报告。 8. 严禁：用而不引、引而不用、伪造数据、抄袭剽窃、买卖论文等学术不端行为。 9. 认真填写毕业设计过程记录，每周至少填写1次，真实记录毕业设计过程。 10. 翻译一篇与毕业设计或专业相关的英文资料或文章，翻译后的中文字数至少在3000字及以上。 |
| 推荐主要参考文献和资料：   1. 陈琳,任芳.基于Python的新浪微博数据爬虫程序设计[J].信息系统工程,2016(09):97-99. 2. Magnus Lie Hetland. Python基础教程(第二版)[M].北京：人民邮电出版社，2014. 3. 朱建平,谢邦昌,骆翔宇,范新妍,曾武雄,郑陈璐.中国房地产网络舆情分析[J].2016, 35(4): 722-741. 4. 杨雪林. 基于大数据的网络舆情监管预测算法研究[J].现代电子技术, 2017,40(24): 28-30. 5. 孙飞显，程世辉，倪天林,靳晓婷.基于新浪微博的负面网络舆情监测研究———针对政府的负面网络舆情研究系列之一[J]. 情报杂志，2015, 34(4) 81-115. 6. 吴宏洲.分词技术的研究与应用——一种快速分词的实现[J].电脑知识与技术,2015,11(06):179-182+186. 7. 叶晓龙.中文分词关键技术研究[J].湖北农机化,2017(06):54-55. 8. 陈昌浩,范太华.改进的HMM模型在特征抽取上的应用[J].计算机测量与控制,2018,26(04):217-220+224. 9. 吴熠潇.中文分词相关算法研究[J].科技经济导刊,2018(02):122-123. 10. 吴立金,简阳,张凯,唐龙利,韩新宇.基于Python语言的GUI自动化测试脚本技术研究[J].计算机测量与控制,2015,23(10):3330-3332+3337. 11. 叶飞.隐马尔可夫模型算法基础探析[J].铜陵学院学报,2014,13(03):108-112. 12. 于长雷,赵红. 重大网络突发事件对房地产市场的影响研究[J].管理评论， 2016，28(8): 66-70. 13. Wenjuan Cui, Pengfei Wang, Yi Du, Xin Chen, Danhuai Guo, Jianhui Li, Yuanchun Zhou. An algorithm for event detection based on social media data[J]. Neurocomputing, 2017, 254: 53–58. 14. Kang Xu, Guilin Qi, Junheng Huang, Tianxing Wu, Xuefeng Fu. Detecting bursts in sentiment-aware topics from social media[J]. Knowledge-Based Systems,2018, 141: 44–54. |

2017年11月15日